



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 9/02, 9/04, 1/20, 15/00, C07H 21/04</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/58649</b> <b>(43) International Publication Date:</b> 18 November 1999 (18.11.99)
<b>(21) International Application Number:</b> PCT/US99/07041 <b>(22) International Filing Date:</b> 13 May 1999 (13.05.99)  <b>(30) Priority Data:</b> 60/085,258      13 May 1998 (13.05.98)      US  <b>(71) Applicant (for all designated States except US):</b> UNIVERSITY OF MARYLAND [US/US]; Office of Technology Liaison, 4312 Knox Road, College Park, MD 20742 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> CUNNINGHAM, Francis, X. [US/US]; 2727 Washington Avenue, Chevy Chase, MD 20815 (US). DELLAPENNA, Dean [US/US]; 4135 Longknife Road, Reno, NV 89557 (US). MOEHS, Charles, P. [US/US]; 1208 Charles Drive, Reno, NV 89509 (US).  <b>(74) Agents:</b> GOLDHUSH, Douglas, H. et al.; Nikaido, Marmelstein, Murray & Oram LLP, Metropolitan Square, Suite 330 - G Street Lobby, 655 Fifteenth Street, N.W., Washington, DC 20005-5701 (US).		<b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> METHODS OF MODIFYING THE PRODUCTION OF ISOPENTENYL PYROPHOSPHATE, DIMETHYLALLYL PYROPHOSPHATE AND/OR ISOPRENOIDS  <b>(57) Abstract</b> <p>The present invention provides methods for enhancing and reducing the levels of IPP, DMAPP and/or isoprenoids in a host cell. Nucleic acid sequences encoding DXP synthase, GAP dehydrogenase, and LYTB as well as vectors containing the same and host cells transformed with the vectors.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS OF MODIFYING THE PRODUCTION OF ISOPENTENYL  
PYROPHOSPHATE, DIMETHYLALLYL PYROPHOSPHATE AND/OR  
ISOPRENOIDS

BACKGROUND OF THE INVENTION

5

Field of the Invention

The present invention is directed to genes encoding deoxyxylulose-5-phosphate synthase (*dxps*), glyceraldehyde-3-phosphate dehydrogenase (*gapd*), and the *lysB* gene product, as well as vectors containing the same and hosts transformed with said vectors. The present invention also provides methods for  
10 modifying the production of isopentenyl pyrophosphate (IPP) and/or dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid compound (i.e., a compound derived from IPP and/or DMAPP). Additionally, the present invention provides a method for screening for procaryotic and eukaryotic genes encoding enzymes that participate in the nonmevalonate pathway leading to IPP and DMAPP.

15

Background of the Invention

A plethora of chemical compounds are produced by what are collectively known as the pathways of isoprenoid (a.k.a. terpene) biosynthesis. The one feature in common to the many isoprenoids (more than 20,000 have been identified in plants; Chappell, 1995) is their biosynthesis from the central metabolites and  
20 building blocks for all isoprenoid compounds: the 5 carbon compounds isopentenyl pyrophosphate (IPP) and its allylic isomer dimethylallyl pyrophosphate (DMAPP). The interconversion of IPP and DMAPP is reversible, and is catalyzed by the enzyme IPP isomerase (IPI). Using the IPP and DMAPP building blocks in various combinations, a modular assembly process that produces compounds of 5, 10, 15,  
25 20 or more carbons (in multiples of 5) allows the biosynthesis of the basic skeletons for the many and various isoprenoids with a relatively small number of basic reaction steps. The C<sub>40</sub> skeleton of plant carotenoid pigments, for instance, is assembled from two molecules of a C<sub>20</sub> compound, geranylgeranyl pyrophosphate (GGPP), that is itself assembled from 3 units of IPP and 1 unit of DMAPP, and that  
30 also serves as a precursor for many other branches of the isoprenoid pathway in plants.

Isoprenoids are formed from IPP and DMAPP in at least three different

compartments of plants cells: the cytosol/endoplasmic reticulum, the mitochondria (and/or Golgi apparatus), and the plastids (McGarvey and Croteau, 1995). The source of IPP and DMAPP for isoprenoid biosynthesis in these different compartments has long been a matter of some controversy and debate (see Bach, 1995; McGarvey and Croteau, 1995). The well-known "classical" or acetate/mevalonate route to IPP and DMAPP proceeds from acetyl-CoA via 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) and mevalonic acid (MVA). The critical, rate-determining and irreversible step in this route in animals is the reduction of HMG-CoA to produce MVA, catalyzed by the enzyme HMG-CoA reductase (HMGR; EC 1.1.1.34; reviewed in Goldstein and Brown, 1990). This enzyme also appears to provide an important control point for substrate flow into certain isoprenoids in plants (Bach, 1986; Chappell et al., 1995; Schaller et al., 1995; Stermer et al., 1994). HMGR has been localized in the cytosol of plants, associated with membranes of the endoplasmic reticulum (reviewed in Bach, 1995). There is no convincing experimental evidence that this key enzyme also resides in the plastids or mitochondria of plants. An uptake by plastids of IPP to support isoprenoid biosynthesis in this organelle has been suggested in the past (Kreuz and Kleinig, 1984). However, it has been difficult to reconcile a cytosolic source of IPP for plastid isoprenoids with the effective inhibition of cytoplasmically-made isoprenoids (e.g. sterols) but not of plastid-made isoprenoids (e.g. carotenoids) or mitochondrial ones (quinones) by mevinolin, a specific inhibitor of HMG-CoA reductase (discussed in Bach, 1995 and McGarvey and Croteau, 1995). Furthermore, the demonstrated ability of isolated chloroplasts of *Acetabularia* (Moore and Sheperd, 1977) and spinach (Schulze-Seibert and Schultz, 1987) to synthesize carotenoids and other isoprenoids from  $^{14}\text{CO}_2$  necessarily implies a chloroplast pathway for IPP biosynthesis. Further confusion arose from reports that radiolabelled acetate is efficiently incorporated by isolated chloroplasts into fatty acids but not into carotenoids (see Kleinig, 1989; Liedvogel, 1986; Schulze-Seibert and Schultz, 1987), a puzzling observation that has been explained by invoking a metabolic "channeling" that somehow renders the newly-formed acetyl-CoA pool available for fatty acid biosynthesis but unavailable for IPP biosynthesis (Kleinig, 1989).

An attractive explanation for these many puzzling and apparently contradictory observations is provided by the suggestion of an "alternative" or non-mevalonate pathway for the biosynthesis of IPP and/or DMAPP in the chloroplasts of plants (Lichtenthaler et al., 1997) and in algae (Schwender et al., 1996).  
5 Existence of a distinct non-mevalonate biochemical pathway to IPP and/or DMAPP, using pyruvate rather than acetate as a substrate, was posited for certain bacteria, including *Escherichia coli*, as early as 1981 (Pandian et al., 1981), but this alternative pathway in bacteria received little attention until relatively recently (Zhou and White, 1991; Rohmer et al., 1993 and Horbach et al., 1993; see Bach, 1995  
10 for a critical discussion of this topic). It appears that many bacteria do not have the acetate/mevalonate pathway for biosynthesis of IPP. For instance, a BLAST search of the *E. coli* genome, the entire sequence for which has now been deposited in GenBank, does not identify any open reading frames with significant similarity to known bacterial, plant or mammalian HMG-CoA reductases. Nor can  
15 we discern an HMG-CoA reductase homologue in the genome sequence data base of the cyanobacterium *Synechocystis* 6803. Because ancestors of the cyanobacteria are the presumptive progenitors of plant chloroplasts, the non-mevalonate route to IPP in these photosynthetic procaryotes is likely to resemble that in plant chloroplasts.

20 Recent reports indicate that plant and algal chloroplast isoprenoids, including carotenoids (Lichtenthaler et al., 1997; Schwender et al., 1996), and the medically-important isoprenoid paclitaxel (a.k.a. TAXOL<sup>®</sup>) (Eisenreich et al., 1996) are made from IPP supplied via a non-mevalonate pathway. Whether this pathway supplies  
25 IPP for carotenoids and other isoprenoids manufactured in the various nonphotosynthetic plant plastids (e.g. chromoplasts, leucoplasts, amyloplasts) or in etioplasts or developing chloroplasts is an open question. Several recent reports describe a thiamin-dependent enzyme from *E. coli* (Sprenger et al., 1997; Lois et al., 1998) and from peppermint (Lange et al., 1998) that converts pyruvate and glyceraldehyde-3-phosphate (GAP) to deoxyxylulose-5-phosphate (DXP). A  
30 homologue of this enzyme was earlier described for *Arabidopsis thaliana* (Mandel et al., 1996), but the function of the *A. thaliana* polypeptide had not been ascertained.

The evidence is compelling that DXP is a substrate in the pathway leading to IPP or DMAPP in plants, algae, and certain bacteria including *E. coli* (Arigoni et al., 1997). However, DXP also is a substrate leading to other essential compounds in these organisms, including thiamin and pyridoxal (see Lois et al., 1998).  
5 Therefore the biochemical reaction catalyzed by the DXP synthase enzyme is not dedicated or restricted to providing substrate for the pathway or pathways leading to IPP and/or DMAPP, and isoprenoids derived therefrom. Because the product of the reaction is shared with several other pathways, the DXP synthase enzyme would not seem to be an obvious candidate for a controlling step in the pathway  
10 leading to isoprenoids. An ability to modify isoprenoid production, enhancing production of desired compounds or reducing that of undesirable compounds, would be quite advantageous in many applications.

Plausible routes to IPP from pyruvate and GAP via DXP have been suggested for plants and bacteria (Schwender et al., 1996; Rohmer et al., 1993 and  
15 1996; see Bach, 1995). A gene encoding what may be the second enzyme in the pathway, DXP reductoisomerase, has recently been identified in *Escherichia coli* (Takahashi et al., 1998). The precise nature and sequence of subsequent biochemical reactions in the nonmevalonate pathway, and whether regulation of flux through this pathway occurs primarily at a single controlling step, as with HMG-  
20 CoA reductase for the mevalonate pathway, are so far unknown. We have taken a genetic approach to delineate the pathway leading to IPP and DMAPP in *E. coli* and plants.

We earlier made the surprising observation that the reversible interconversion of IPP and DMAPP limits the accumulation of carotenoids in *E. coli*  
25 (an observation alluded to in Sun et al., 1996 and Sun et al., 1998). Others (Kajiwara et al., 1997) have made a similar observation. As a direct result of continuing to pursue the same empirical approach (using a colored isoprenoid compound to "report" on the amount of IPP and DMAPP available for isoprenoid biosynthesis) which revealed that the isomerization of IPP is a limiting step in  
30 isoprenoid biosynthesis (see U.S. Patent No. 5,744,341), we have now identified several other genes which can be used to modify IPP, DMAPP and/or isoprenoid production in host cells.

For two of the these genes the functions of the products, glyceraldehyde-3-phosphate dehydrogenase and DXP synthase, are well known. The role of the product of the third, the *lytB* gene, is still obscure. Because certain mutations in this gene in *Escherichia coli* confer an ability to tolerate elevated levels of penicillin (Harkness et al., 1992; Gustafson et al., 1993) the *lytB* gene product would seem to play some role in or otherwise influence the biosynthesis of peptidoglycan. It has been suggested that the *lytB* gene product is not an enzyme directly involved in the pathway leading to peptidoglycan but rather exerts some control on the production of the global regulator molecule guanosine 3',5'-bisphosphate (ppGpp) (Gustafson et al., 1993; Rodionov and Ishiguro, 1995). The evidence in support of this suggestion is correlative and indirect and therefore by no means convincing. Because *lytB* influences the accumulation of certain isoprenoids in *E. coli* (see below), because an analysis of the nearly twenty completed bacterial genomes reveals it to be one of the very few genes found only in those bacteria that also contain genes encoding DXP synthase and DXP reductoisomerase, and because we have isolated an apparently chloroplast-targeted *lytB* cDNA from a eucaryotic plant, we speculate that *lytB* may encode an enzyme that catalyzes one of the later reactions in the nonmevalonate pathway leading to IPP and DMAPP in certain bacteria and in plant chloroplasts. To avoid confusion, it is important to note that another gene, which is completely unrelated to the gene under consideration here, has unfortunately also been given the name *lytB* (see, e.g., Garcia et al., 1999).

There remains a need for methods to enhance the production of desirable isoprenoids and reduce the accumulation of undesirable isoprenoids in plants, fungi and bacteria that contain a nonmevalonate pathway. There also remains a need in the art for methods for screening for procaryotic and eukaryotic genes encoding enzymes of isoprenoid biosynthesis and metabolism.

#### SUMMARY OF THE INVENTION

A subject of the present invention is an isolated nucleic acid sequence which encodes for a protein having DXP synthase enzyme activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 1 or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least

85% identical to SEQ ID NO: 2.

Another subject of the present invention is an isolated nucleic acid sequence which encodes for a protein having GAP dehydrogenase enzyme activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 3 or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 85% identical to SEQ ID NO: 4.

A further subject of the present invention is an isolated nucleic acid sequence which encodes for a protein having LYTb activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 5 or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 85% identical to SEQ ID NO: 6.

The invention also includes vectors which comprise any of the nucleic acid sequences listed above, and host cells transformed with such vectors.

Another subject of the present invention is a method of enhancing the production of IPP, DMAPP and/or an isoprenoid compound in a host cell, comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having DXP synthase, GAP dehydrogenase and/or LYTb activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence, thereby producing the protein.

Yet another subject of the present invention is a method of modifying the production of IPP, DMAPP and/or an isoprenoid compound in a host cell, comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein which modifies DXP synthase, GAP dehydrogenase and/or LYTb activity in the host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence, thereby producing the protein.

The present invention also includes a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having DXP synthase, GAP dehydrogenase and/or LYTb activity, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a



promoter; and expressing the heterologous nucleic acid sequence.

Also included is a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein which modifies DXP synthase, GAP dehydrogenase and/or LYT activity in the host cell, the method comprising  
5 inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence.

Another subject of the present invention is to provide a method for screening for eukaryotic genes which encode enzymes involved in isoprenoid biosynthesis  
10 and metabolism.

### BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following detailed description when considered in  
15 connection with the accompanying drawings, wherein:

Figure 1 is an illustration showing that IPP and DMAPP serve as the central metabolites leading to an immense variety of different isoprenoid compounds in plants, bacteria, and other organisms. The mevalonic acid pathway via acetyl-CoA in the cytosol of plants and in animals is well characterized. A second route from  
20 pyruvate and GAP, not yet elucidated, is thought to operate in the chloroplasts of plants and algae, in cyanobacteria, and in many bacteria. Later reaction steps in the pathway remain to be determined and are therefore denoted with question marks. The pathway is shown leading to a box containing IPP and DMAPP because it has not been established which of these two is the initial product of this  
25 pathway. Abbreviations: G3P, glyceraldehyde-3-phosphate; G3PD, GAP dehydrogenase; DXPS, DXP synthase; IPP, isopentenyl pyrophosphate; IPI, IPP isomerase; DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GPP geranyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate.

Figure 2 illustrates a route to IPP/DMAPP from pyruvate and glyceraldehyde-3-

phosphate (GAP), which is thought to operate in the chloroplasts of plants and algae, in cyanobacteria, and in many bacteria including *E. coli*. Later reaction steps in the pathway remain to be determined and are therefore denoted with question marks. The pathway is shown leading to a box containing IPP and DMAPP because it has not been established which of these two is the initial product of this pathway. Enzymes of interest are shown in bold white text in a black box. Abbreviations: DMAPP, dimethylallyl pyrophosphate; DXPS, deoxyxylulose-5-phosphate synthase; DXPR, deoxyxylulose-5-phosphate reductoisomerase; FPS, farnesyl pyrophosphate synthase, GAPD, glyceraldehyde-3-phosphate dehydrogenase; GGPP, geranylgeranyl pyrophosphate; IPI, isopentenyl pyrophosphate isomerase; IPP, isopentenyl pyrophosphate.

Figure 3 illustrates that the  $C_{40}$  carotenoid phytoene is derived by a head-to-head condensation of two molecules of the  $C_{20}$  GGPP compound, which itself is assembled from 3 molecules of IPP and 1 molecule of DMAPP. Abbreviations: FPP, farnesyl pyrophosphate; GPP geranyl pyrophosphate; PPPP, prephytoene pyrophosphate.

Figure 4 is a schematic illustration and restriction map of plasmid pAC-LYC (Cunningham et al., 1994) which contains genes (*crtE*, *crtB*, and *crtI*) of *Erwinia herbicola* encoding all of the enzymes required for production of the pink-colored isoprenoid pigment lycopene from the colorless IPP and DMAPP compounds. Cm, chloramphenicol resistance gene.

Figure 5A is a cDNA sequence and Figure 5B is the predicted amino acid sequence of a putative DXP synthase isolated from a flower cDNA library of *Tagetes erecta* (SEQ ID NOS: 1 and 2). The cDNA is incorporated into the plasmid pMarDXPS.

Figure 6A is a cDNA sequence and Figure 6B is the predicted amino acid sequence of a chloroplast isoform of GAP dehydrogenase isolated from *Arabidopsis thaliana* (SEQ ID NOS: 3 and 4). The cDNA is incorporated into the plasmid pAtG3PD.

Figure 7A is a cDNA sequence and Figure 7B is the predicted amino acid sequence of a LYTB protein isolated from an *Adonis palaestina* flower cDNA library (SEQ ID NOS: 5 and 6). The sequence is of clone Ipi3 except that the 14 bp at the N-terminus were obtained from the slightly longer cDNA clone Ipi18. The cDNA is incorporated into the plasmid pApLYTB.

Figure 8 is an alignment of the predicted amino acid sequences of LYTB from *Adonis palaestina*, *Synechocystis PCC6803* and *E. coli*. The N-terminal extension of the *Adonis* polypeptide, relative to that of *Synechocystis PCC6803*, is predicted by the program ChloroP (Emanuelsson et al., 1999) to constitute a chloroplast transit peptide, serving to target the polypeptide to this organelle in plants. Black boxes with white letters are used where all three of the aligned residues are identical. Grey boxes with black letters are used where two of the three aligned residues are identical. Abbreviations: Ap, *Adonis palaestina*; Sy, *Synechocystis PCC6803*; Ec, *Escherichia coli*.

Figure 9 is a schematic representation of the mapping of an *E. coli* genomic fragment to ascertain which of the genes in this fragment will enhance or impair lycopene accumulation in *E. coli*. Deletion mapping of an *E. coli* genomic fragment (the insert in plasmid pEc3.9 is essentially identical to GenBank U32768: 4879..8819) with genes encoding DXP synthase (orf620), FPP synthase (*ispA*) and the small subunit of exonuclease VII (*xseB*). Numbers to the right indicate relative lycopene accumulation per mL of liquid culture with plasmids in lycopene-accumulating *E. coli* strain TOP10. The insert in pEc3.9 is oriented in the forward direction in the multicopy plasmid vector pBluescript SK-. The *EcoRI* and *SmaI* sites in the vector preceding the genomic fragment and a *KpnI* site following it were used, along with *NdeI*, *SmaI*, and *SalI* sites in the genomic DNA, to construct the deletion subclones illustrated. Incomplete genes and open reading frames are not shown.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

The term "isoprenoid" is intended to mean any member of the class of

naturally occurring compounds whose carbon skeletons are composed, in part or entirely of isopentyl C<sub>5</sub> units. Preferably, the carbon skeleton is of an essential oil, a fragrance, a rubber, a carotenoid, or a therapeutic compound, such as paclitaxel.

A search of the public databases (GenBank) for genes encoding  
5 homologues of the *E. coli* *lytB* gene product unearthed such genes in twelve bacteria and in the cyanobacterium *Synechocystis* PCC6803. A recent publication (Potter et al., 1998) discusses the distribution of this gene in various bacteria. Our description herein of a homologue in the plant *Adonis palaestina* reveals for the first time that this gene is present and expressed in a eucaryotic organism. A recently  
10 deposited genomic DNA sequence for the green plant *Arabidopsis thaliana* (GenBank accession number AL035521) contains what appears to be a gene (the probable coding sequence is interrupted by several apparent introns) encoding LYTB in this organism. The predicted sequence of this *Arabidopsis* LYTB is somewhat uncertain (a comparison of the *Adonis* sequence with that listed given  
15 in AL035521 for *Arabidopsis* suggests that several of the exon-intron junctions predicted for this gene in the GenBank record are incorrect), but sequence identity in a comparison with the *Adonis* sequence is ca. 80% or more. Partial cDNA sequences in the data base of expressed sequence tags (dbEST) that predict peptides with sequence similarity to portions of the *Adonis* LYTB sequence indicate  
20 that homologues exist and mRNAs encoding LYTB are produced in several other plant species including rice (D45948), loblolly pine (AA556723) and soybean (AI437981). Both the *Adonis* and *Arabidopsis* predicted amino acid sequences are more than 60% identical to that predicted by the cyanobacterium *Synechocystis* PCC6803 gene, and the two plant and the cyanobacterial sequences are more than  
25 30% identical to the predicted *E. coli* gene product. An alignment of the *Adonis*, *Synechocystis* PCC6803 and *E. coli* predicted amino acid sequences is shown in Figure 7. A number of regions and residues conserved in LYTB are indicated in this Figure.

The term "LYTB activity" is intended to mean the ability of LYTB to affect the  
30 production of IPP, DMAPP and/or isoprenoids in a host cell containing the *lytB* gene or DNA copy of the *lytB* mRNA. It has not yet been confirmed that the LYTB protein is, in fact, an enzyme. The precise role of the LYTB protein in affecting the

production of isoprenoids has not been established. However, whatever the mechanism of action of the *lytB* gene product, we intend to cover that mechanism of action by the term "LYTB activity".

5 The present invention is directed to an isolated nucleic acid sequence which encodes for a protein having DXP synthase enzyme activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 1 or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 85% identical to SEQ ID NO: 2. Preferably, the nucleic acid sequence is at least 90%, at least 95% or completely identical to SEQ ID NO: 1, or the nucleic acid  
10 sequence encodes a protein which has an amino acid sequence which is at least 90%, at least 95% or completely identical to SEQ ID NO: 2.

Another subject of the present invention is an isolated nucleic acid sequence which encodes for a protein having GAP dehydrogenase enzyme activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 3 or the nucleic  
15 acid sequence encodes a protein which has an amino acid sequence which is at least 85% identical to SEQ ID NO: 4. Preferably, the nucleic acid sequence is at least 90%, at least 95% or completely identical to SEQ ID NO: 3, or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 90%, at least 95% or completely identical to SEQ ID NO: 4.

20 A further subject of the present invention is an isolated nucleic acid sequence which encodes for a protein having LYTB activity, wherein the nucleic acid sequence is at least 85% identical to SEQ ID NO: 5 or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 85% identical to SEQ ID NO: 6. Preferably, the nucleic acid sequence is at least  
25 90%, at least 95% or completely identical to SEQ ID NO: 5, or the nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 90%, at least 95% or completely identical to SEQ ID NO: 6.

In each case, sequence similarity is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics  
30 Computer Group (University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705), MEGAlign (DNASar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), or MacVector (Oxford Molecular Group, 2105 S.

Bascom Avenue, Suite 200, Campbell, California 95008). Such software matches similar sequences by assigning degrees of identity to various substitutions, deletions, and other modifications. Conservative (i.e. similar) substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (see Kyte and Doolittle, *J. Mol. Biol.* **157**: 105-132 (1982)), or on the basis of the ability to assume similar polypeptide secondary structure (see Chou and Fasman, *Adv. Enzymol.* **47**: 45-148 (1978)). If comparison is made between nucleotide sequences, preferably the length of comparison sequences is at least 50 nucleotides, more preferably at least 60 nucleotides, at least 75 nucleotides or at least 100 nucleotides. It is most preferred if comparison is made between the nucleic acid sequences encoding the protein coding regions necessary for protein activity. If comparison is made between amino acid sequences, preferably the length of comparison is at least 20 amino acids, more preferably at least 30 amino acids, at least 40 amino acids or at least 50 amino acids. It is most preferred if comparison is made between the amino acid sequences in the protein coding regions necessary for protein activity.

The present inventors have isolated eukaryotic genes encoding DXP synthase from *Tagetes erecta* (marigold), GAP dehydrogenase from *Arabidopsis thaliana*, and LYTB from *Adonis palaestina*. All were identified on the basis of an enhancement of lycopene accumulation in *E. coli*. The *E. coli* DXP synthase was also identified in this same way.

Suitable vectors according to the present invention comprise a gene encoding one or more of the above-identified enzymes involved in IPP, DMAPP and/or isoprenoid biosynthesis or metabolism, wherein the gene is operably linked to a suitable promoter. Suitable promoters for the vector can be constructed using techniques well known in the art (see, for example, Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing and Wiley Interscience, New York, 1991). Suitable vectors for

eukaryotic expression in plants are described in Frey et al., *Plant J.* (1995) 8(5):693 and Misawa et al, 1994a. Suitable vectors for prokaryotic expression include pACYC184, pUC119, and pBR322 (available from New England BioLabs, Beverly, MA) and pTrcHis (Invitrogen) and pET28 (Novagen) and derivatives thereof. The  
5 vectors of the present invention can additionally contain regulatory elements such as promoters, repressors, selectable markers such as antibiotic resistance genes, etc., the construction of which is very well known in the art.

One or more of the genes encoding the enzymes as described above, when cloned alone or in combination into a suitable expression vector, can be used to  
10 overexpress these enzymes in a plant expression system or to inhibit the expression of these enzymes. For example, a vector containing one or more of the genes of the invention may be used to increase the amount of isoprenoids in an organism and thereby alter the nutritional or commercial value or pharmacology of the organism.

Therefore, the present invention includes a method of enhancing the production of IPP, DMAPP and/or an isoprenoid in a host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having DXP synthase, GAP dehydrogenase and/or LYTB activity, wherein the  
15 heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence.

The invention also includes a method of modifying the production of IPP, DMAPP and/or an isoprenoid in a host cell, the method comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which  
20 encodes for a protein which modifies DXP synthase, GAP dehydrogenase and/or LYTB activity in the host cell, relative to an untransformed host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence.

The invention further includes a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having DXP synthase, GAP dehydrogenase and/or LYTB activity, the method comprising  
30 inserting into the host cell a vector comprising the heterologous nucleic acid

sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

The invention also includes a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein which modifies  
5 DXP synthase, GAP dehydrogenase and/or LYT<sub>B</sub> activity in the host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

10 Preferably, the isoprenoid comprises a compound derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP). By the term "derived from", we mean that the isoprenoid contains at least one of the listed compounds as a building block. It is also preferable if the isoprenoid comprises at  
15 least one member selected from the group consisting of a diterpene, a carotenoid, an essential oil, a fragrance, an isoprene, a cytokinin, a rubber, a quinone, a sterol, a hopanoid, a triterpene, a steroid, a prenylated protein, a phytoalexin, a gibberellin, a tocopherol, a dolichol, a chlorophyll and a therapeutic compound.

It is most preferred if the isoprenoid comprises at least one member selected  
20 from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound, such as paclitaxel.

The heterologous nucleic acid sequence may originate from a eukaryotic or procaryotic cell. By the term "originate from", we intend to mean that the sequence information for the heterologous nucleic acid came from the eukaryotic or the  
25 procaryotic cell. However, the specific nucleic acid itself does not have to be from the organism. The nucleic acid may come from the organism, or it may be synthetically produced using recombinant nucleic acid techniques known in the art.

Preferably, the heterologous nucleic acid sequence comprises a nucleotide sequence for *dxps*, *gapd* and/or *lytB*. It is most preferred that the heterologous  
30 nucleic acid sequence comprises a nucleotide sequence which encodes a *dxps*, *gapd* and/or a *lytB* gene and is at least 85% identical, preferably at least 90%, at least 95% or completely identical, to SEQ ID NO: 1, 3 and/or 5, respectively, or the



nucleic acid sequence encodes a protein which has an amino acid sequence which is at least 90%, at least 95% or completely identical to SEQ ID NO: 2, 4 and/or 6, respectively. Identity is determined as noted above.

5 The term "modifying the production" in the methods of the invention means that the amount of target compounds produced (e.g., IPP, DMAPP and/or isoprenoids) can be enhanced or reduced, as compared to an untransformed host cell. Thus, in accordance with an embodiment of the present invention, the production or the biochemical activity of the target compounds (or the enzymes which catalyze their formation) may be reduced or inhibited by a number of different  
10 approaches available to those skilled in the art, including but not limited to such methodologies or approaches as anti-sense (e.g., Gray et al., 1992), ribozymes (e.g., Wegener et al., 1994), co-suppression (e.g. Fray and Grierson, 1993), targeted disruption of the gene (e.g., Schaefer et al., 1997), intracellular antibodies (e.g., see Rondon and Marasco, 1997) or whatever other approaches rely on the  
15 knowledge or availability of the nucleic acid sequences of the invention, or the proteins encoded thereby.

Host systems according to the present invention can comprise any organism that utilizes a nonmevalonate (i.e., via DXP) pathway for production of IPP and/or DMAPP. Organisms which produce isoprenoids using IPP and/or DMAPP derived  
20 from a nonmevalonate pathway include plants, algae, certain bacteria, cyanobacteria and other photosynthetic bacteria. Transformation of these hosts with vectors according to the present invention can be done using standard techniques. See, for example, Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; Ausubel  
25 et al., Current Protocols in Molecular Biology, Greene Publishing and Wiley Interscience, New York, 1991.

Alternatively, transgenic organisms can be constructed which include the nucleic acid sequences of the present invention. The incorporation of these sequences can allow the controlling of isoprenoid biosynthesis, content, or  
30 composition in the host cell. These transgenic systems can be constructed to incorporate sequences which allow for the overexpression of the various nucleic acid sequences of the present invention. Transgenic systems can also be

constructed which allow for the underexpression of the various nucleic acid sequences of the present invention. Such systems may contain anti-sense expression of the nucleic acid sequences of the present invention. Such anti-sense expression would result in the accumulation of the substrates of the enzyme encoded by the sense strand.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

### EXAMPLE 1

#### Isolation of plant cDNAs and bacterial genes that enhance carotenoid accumulation in *Escherichia coli*

The plasmid pAC-LYC (Cunningham et al., 1994 and 1996) contains genes encoding all of the enzymes required for the formation of lycopene from IPP and DMAPP (see Figure 3). Cells of *E. coli* containing the plasmid pAC-LYC accumulate the carotenoid lycopene and thereby form colonies on solid growth medium that are pink in color (Cunningham et al., 1994). When cDNA or genomic libraries of plants, cyanobacteria, or bacteria are introduced into the lycopene-accumulating *E. coli* strain (the complete methodology is contained in Cunningham et al., 1994 and 1996), a rare few of the colonies formed after spreading of the cell culture onto solid growth medium are much deeper in color than is typical of the vast majority of the colonies. The enhancement of color and concomitant carotenoid pigment accumulation in *E. coli* is *prima facie* evidence for an enhancement of flux into and through the pathway or pathways leading to the immediate precursors of carotenoids and all other isoprenoid compounds: IPP and its allylic isomer DMAPP. Genes encoding subsequent enzymes in the pathway leading to lycopene are already present in the multicopy plasmid pAC-LYC (see Figure 3).

Among the cDNAs and genes isolated with this screening methodology are: *Tagetes erecta* (marigold) cDNAs encoding a homologue of DXP synthase (the DNA sequence of the longest is given in Figure 4), an *E. coli* genomic clone containing the DXP synthase (and no other complete open reading; a genomic

fragment from the *SphI* site which lies at base 5210 in the sequence listed under GenBank U32768 to the *NdeI* site at base 8003 in this GenBank sequence was found to enhance carotenoid accumulation in *E. coli*. This region encompasses the open reading frame earlier referred to as f620, and now as the DXP synthase),  
5 *Arabidopsis thaliana* cDNAs encoding a chloroplast isoform (Shih *et al.*, 1991 and 1992) of the enzyme GAP dehydrogenase (the sequence of the longest of these isolated cDNAs is given in Figure 5), and an *Adonis palaestina* cDNA encoding a homologue of the *E. coli* *lytB* gene product (Figure 6). The enhancement of isoprenoid biosynthesis in *E. coli* (as indicated by lycopene accumulation) when  
10 cDNAs or genes encoding DXP synthase or GAP dehydrogenase are introduced can be understood in the context of the biochemical pathway that has been postulated for production of IPP/DMAPP in *E. coli* (and in other bacteria, cyanobacteria and plants). However, at the date of the provisional filing of this application, there had been no indications that increased flux through the pathway  
15 to IPP/DMAPP might be obtained by increasing the expression of either of these enzymes. A publication after the priority date of this application (Harker and Bramley (1999)) shows a salutary effect on isoprenoid accumulation in *E. coli* in which foreign DXP synthase genes are introduced.

The mechanism of enhancement of isoprenoid production by introduction of  
20 cDNAs encoding LYT B is not yet known. The nonmevalonate pathway leading to IPP and/or DMAPP has not yet been elucidated and these cDNAs may encode an enzyme subsequent to DXP synthase in the pathway. Alternatively, the enhancement of isoprenoid accumulation may involve a mechanism less direct (e.g., as for GAP dehydrogenase, involvement in biochemical reactions that utilize  
25 or supply the substrates GAP and pyruvate, or by exerting a regulatory influence on isoprenoid pathways).

## EXAMPLE 2

### *Arabidopsis thaliana*, *Tagetes erecta*, and *Adonis palaestina* cDNA Libraries

A size-fractionated 2-3 kbp cDNA library of *A. thaliana* in lambda ZAPII  
30 (Kieber *et al.*, 1993) was obtained from the Arabidopsis Biological Resource Center at The Ohio State University (stock number CD4-15). Other size fractionated

libraries were also obtained (stock numbers CD4-13, CD4-14, and CD4-16). The cDNA libraries for *Tagetes erecta* (marigold) and *Adonis palaestina* (pheasant's eye) were constructed for us by Stratagene using cDNAs isolated from flower tissues. An aliquot of each library was treated to cause a mass excision of the cDNAs and thereby produce a phagemid library according to the instructions provided by the supplier of the cloning vector (Stratagene; *E. coli* strain XL1-Blue and the helper phage R408 were used). The titre of the excised phagemid was determined and the library was introduced into a lycopene-accumulating strain of *E. coli* TOP10 F' (this strain contained the plasmid pAC-LYC) by incubation of the phagemid with the *E. coli* cells for 15 min at 37°C. Cells had been grown overnight at 30°C in LB medium supplemented with 2% (w/v) maltose and 10 mM MgSO<sub>4</sub> (final concentration), and harvested in 1.5 ml microfuge tubes at a setting of 3 on an Eppendorf microfuge (5415C) for 10 min. The pellets were resuspended in 10 mM MgSO<sub>4</sub> to a volume equal to one-half that of the initial culture volume. Transformants were spread on large (150 mm diameter) LB agar petri plates containing antibiotics to provide for selection of cDNA clones (ampicillin) and maintenance of pAC-LYC (chloramphenicol). Approximately 10,000 colony forming units were spread on each plate. Petri plates were incubated at room temperature for 2 to 7 days to allow maximum color development. Plates were screened visually with the aid of an illuminated 3x magnifier and a low power stage-dissecting microscope for the rare deep pink colonies that could be observed in the background of paler pink colonies.

### EXAMPLE 3

#### Enhancement of Carotenoid Accumulation in *E. coli* by cDNAs and genes encoding LYTB, DXP synthase, IPP isomerase and GAP Dehydrogenase Individually and in Combination.

Attempts were made to maximize production of three very different carotenoids in *E. coli* engineered to accumulate these isoprenoid pigments. The specific carotenoids chosen were the linear and lipophilic pink compound lycopene, the bicyclic hydrocarbon  $\beta$ -carotene and the much more polar dihydroxy, bicyclic carotenoid zeaxanthin. DXP synthase, IPP isomerase, LYTB, and GAP

dehydrogenase had been observed to enhance the pigmentation of *E. coli* colonies growing on solid agar medium. For a more quantitative appraisal of the influence of these genes and cDNAs on carotenoid accumulation, the amounts of carotenoid pigments in liquid cultures were examined with the results given in Figure 1. The marigold DXP synthase, an Arabidopsis IPP isomerase and the Adonis LYT B cDNAs gave rise to significant and substantial increases in the amount of carotenoids accumulated per volume of culture. A partially-sequenced Arabidopsis cDNA (GenBank AA605545) with sequence similarity to the Adonis *lytB* cDNA gave results similar to that of the Adonis cDNA (not shown). A *synechocystis* PCC6803 *lytB* gene was less effective but gave a significant enhancement as well. See Table 1. GAP dehydrogenase was slightly detrimental to pigment accumulation in liquid culture, in contrast to earlier observed enhancement for cultures grown on solid media. The influence of this gene on pigment accumulation may depend on the specific growth regimen. A similar result, enhancement for cultures on solid media but reduction in liquid culture, was also obtained for an *E. coli* genomic fragment containing DXP synthase (see Example 4 below and Figure 8). A combination of DXP synthase with IPP isomerase was significantly more effective at enhancing pigment accumulation than was either of the individual cDNAs.

#### EXAMPLE 4

##### Enhancement and Reduction of Lycopene Accumulation in *E. coli* containing multiple copies of genes encoding DXP synthase and Farnesyl Pyrophosphate synthase, respectively

A number of the inserts in plasmids obtained from dark pink colonies selected in screens of the *E. coli* genomic library (see example 1) contained the gene encoding DXP synthase in this organism (see Sprenger et al., 1997; Lois et al., 1998). Deletion mapping of one such *E. coli* genomic fragment indicated that the dark pink colony phenotype and enhanced carotenoid accumulation in liquid culture was conferred by a fragment containing the DXP synthase and no other complete open reading (Figure 8). A genomic fragment containing the *ispA* and *xseB* genes immediately upstream of *dxps* was, in contrast, quite inhibitory for carotenoid accumulation (Figure 8). This diminution in pigmentation results, we

surmise, from the activity of the farnesyl pyrophosphate (FPP) synthase encoded by *ispA* (Fujisaki et al., 1990). The FPP synthase would be expected to compete with geranylgeranyl pyrophosphate (GGPP) synthase (Figure 1) for IPP and DMAPP and thereby deplete the substrate available for carotenoid biosynthesis.

Table 1 illustrates the influence of cDNAs encoding DXP synthase, GAP dehydrogenase, LYT<sub>B</sub>, and IPP isomerase, individually and in combination, on the accumulation of lycopene,  $\beta$ -carotene, or zeaxanthin in three strains of *E. coli* engineered to produce these isoprenoid compounds.

Table 1. Enhancement of carotenoid accumulation in lycopene,  $\beta$ -carotene, and zeaxanthin-accumulating strains of *E. coli* by introduction of plant cDNAs and cyanobacterial genes individually and in combination.

Plasmid	cDNA or Gene Product	Lycopene % of control	$\beta$ -Carotene % of control	Zeaxanthin % of control
pBluescript SK-	Control (empty vector)	(100)	(100)	(100)
pAtipITrc	<i>Arabidopsis thaliana</i> IPI	202 $\pm$ 12	227 $\pm$ 3	223 $\pm$ 4
pTedxps	<i>Tagetes erecta</i> DXPS	203 $\pm$ 11	193 $\pm$ 11	228 $\pm$ 40
pAdlytB	<i>Adonis palaeestina</i> LYT <sub>B</sub>	157 $\pm$ 9	180 $\pm$ 3	168 $\pm$ 2
p6803lytB	<i>Synechocystis</i> PCC 6803 LYT <sub>B</sub>	116 $\pm$ 11	111 $\pm$ 1	109 $\pm$ 2
pAtgapA	<i>Arabidopsis thaliana</i> GAPD	79 $\pm$ 6	nd	79 $\pm$ 3

Data were obtained with *E. coli* strain TOP10 containing pAC-LYC, pAC-BETA or pAC-ZEAX (Cunningham et al., 1994 and 1996, Sun et al., 1996). Carotenoid content (per volume of culture) was measured after 48-72 h growth in LB medium. Cultures were inoculated with freshly grown individual colonies from agar plates. Values are mean  $\pm$  SD for 4 to 13 individual cultures. The empty plasmid pBluescript SK<sup>-</sup> (from Stratagene) served as the control and was used as the vector for all of the plant cDNAs and the *Synechocystis lytB* gene. The influence of GAPD on accumulation of  $\beta$ -carotene was not determined (nd).

## REFERENCES

Each reference cited in this application and/or listed below is hereby incorporated by reference.

Arigoni D, Sagner S, Latzel C, Eisenreich W, Bacher A, Zenk MH (1997) Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal

rearrangement. *Proc. Nat. Acad. Sci USA* 94:10600-110605

Bach TJ. 1986. Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol synthesis? *Lipids* 21:82-88

Bach TJ. 1995. Some new aspects of isoprenoid biosynthesis in plants--a review. *Lipids* 30:191-202

Chappell J. 1995. Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 46:521-47

Chappell J, Wolf F, Proulx J, Cuellar R, Saunders C. 1995. Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl coenzyme A reductase a rate-limiting step for isoprenoid biosynthesis in plants? *Plant Physiol* 109:1337-1343

Cunningham FX Jr, Pogson B, Sun Z, McDonald KA, DellaPenna D, Gantt E. 1996. Functional analysis of the b and e lycopene cyclase enzymes of *Arabidopsis* reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* 8:1613-26

Cunningham FX Jr, Sun, Z, Chamovitz D, Hirschberg J, Gantt E. 1994. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. *Plant Cell* 6:1107-21.

Eisenreich W, Menhard B, Hylands PJ, Zenk MH, Bacher A. 1996. Studies on the biosynthesis of taxol: the taxane skeleton is not of mevalonoid origin. *Proc. Nat. Acad. Sci. U.S.A.* 93:6431-6436

Emanuelsson O, Nielsen H, von Heijne G (1998) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, in press

Fray RG, Grierson D (1993) Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol Biol* 22:589-602

Fujisaki S, Hara H, Nishimura Y, Horiuchi K, Nishino T (1990) Cloning and nucleotide sequence of the ispA gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*. *J Biochem (Tokyo)* 108:995-1000

Garcia P, Gonzalez MP, Garcia E, Lopez R, Garcia JL. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol* 31(4):1275-7

Goldstein JL, Brown MS. 1990. Regulation of the mevalonate pathway. *Nature*

343:425-430

- Gray J, Picton S, Shabbeer J, Schuch W, Grierson D (1992) Molecular biology of fruit ripening and its manipulation with antisense genes. *Plant Mol Biol* 19:69-87
- Gustafson CE, Kaul S, Ishiguro EE. (1993) Identification of the *Escherichia coli* *lytB* gene, which is involved in penicillin tolerance and control of the stringent response. *J Bacteriol* 175:1203-5
- Harker M, Bramley PM. (1999) Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. *FEBS Lett* 448:115-9
- Harkness RE, Kusser W, Qi BJ, Ishiguro EE. (1992) Genetic mapping of the *lytA* and *lytB* loci of *Escherichia coli*, which are involved in penicillin tolerance and control of the stringent response. *Can J Microbiol* 38:975-8
- Horbach S, Sahm H, Welle R. 1993. Isoprenoid biosynthesis in bacteria: two different pathways? *FEMS Microbiol Lett* 111:135-140
- Kajiwara S, Fraser PD, Kondo K, Misawa N (1997) Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochem J* 324:421-426
- Kreuz K, Kleinig H. 1984. Synthesis of prenyl lipids in cells of spinach leaf. Compartmentation of enzymes for formation of isopentenyl diphosphate. *Eur J Biochem* 141:531-535
- Liedvogel B. Acetyl coenzyme A and isopentenyl-pyrophosphate as lipid precursors in plant cells - biosynthesis and compartmentation. *J Plant Physiol* 124:211-222
- Kleinig H. 1989. The role of plastids in isoprenoid biosynthesis. *Ann Rev Plant Physiol Plant Mol Biol* 40:39-59
- Kreuz K, Kleinig H. 1984. Synthesis of prenyl lipids in cells of spinach leaf. Compartmentation of enzymes for formation of isopentenyl diphosphate. *Eur J Biochem* 141:531-535
- Lange BM, Wildung MR, McCaskill D, Croteau R (1998) A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Nat. Acad. Sci USA* 95:2100-2104
- Lichtenthaler HK, Schwender J, Disch A, Rohmer M. 1997. Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent



pathway. FEBS Lett 1997 400:271-274

Lois LM, Campos N, Putra SR, Danielsen K, Rohmer M, Boronat A (1998) Cloning and characterization of a gene from *Escherichia coli* encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. Proc. Nat. Acad. Sci USA 95:2105-2110

Mandel AA, Feldmann K, Herrera-Estrella L, Rocha-Sosa M, Leon P, (1996) CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. Plant J. 9:649-658

McGarvey DJ, Croteau R. 1995. Terpenoid metabolism. Plant Cell 7:1025-1026

Moore FD, Shephard DC. 1977. Biosynthesis in isolated *Acetabularia* chloroplasts. II. Plastid pigments. Protoplasma 92:167-175

Pandian S, Saengchjan S, Raman TS. 1981. An alternative pathway for the biosynthesis of isoprenoid compounds in bacteria. Biochem J 15:675-681

Rodionov DG, Ishiguro EE. (1995) Direct correlation between overproduction of guanosine 3',5'-bisphosphate (ppGpp) and penicillin tolerance in *Escherichia coli*. J Bacteriol 177:4224-4229

Potter S, Yang X, Boulanger MJ, Ishiguro EE. (1998) Occurrence of homologs of the *Escherichia coli* *lytB* gene in gram-negative bacterial species. J Bacteriol 180:1959-61

Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. 1993. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J 295:517-524

Rohmer M, Seemann M, Horbach S, Bringer-Meyer S, Sahm H. 1996. Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. J. Am. Chem. Soc. 118:2564-2566

Rondon IJ, Marasco WA (1997) Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. Annu Rev Microbiol 51:257-283

Sambrook et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss *Physcomitrella*

- patens. Plant J 11:1195-1206
- Schaller H, Grausem B, Benveniste P, Chye ML, Tan CT, Song YH, Chua NH. 1995. Expression of the *Hevea brasiliensis* (H.B.K.) Müll. Arg. 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 in tobacco results in sterol overproduction. Plant Physiol 109:761-770
- Schulze-Seibert D, Schulz G. 1987. b-Carotene synthesis in isolated spinach chloroplasts. Its tight linkage to photosynthetic carbon metabolism. Plant Physiol 84:1233-1237
- Shih MC, Heinrich P, Goodman HM (1991) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate-dehydrogenase from *Arabidopsis thaliana*. Gene 104:133-138
- Shih MC, Heinrich P, Goodman HM (1992) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate-dehydrogenase from *Arabidopsis thaliana*. Erratum. Gene 119:317-319
- Schwender J, Seemann M, Lichtenthaler HK, Rohmer M. 1996. Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*. Biochem J 316:73-80
- Sprenger GA, Schörken U, Wiegert T, Grolle S, de Graaf AA, Taylor, SV, Begley TP, Bringer-Meyer S, Sahm H (1997) Identification of a thiamin-dependent synthase in *Escherichia coli* required for formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. Proc. Nat. Acad. Sci USA 94:12857-12862
- Stermer BA, Bianchini GM, Korth KL. 1994. Regulation of HMG-CoA reductase activity in plants. J. Lipid Research 35:1133-1140
- Sun Z, Gantt E, Cunningham FX Jr (1996) Cloning and functional analysis of the  $\beta$ -carotene hydroxylase of *Arabidopsis thaliana*. J Biol Chem 271: 24349-24352
- Sun Z, Cunningham FX Jr, Gantt E (1998) Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. Proc Nat Acad Sci USA 95: 11482-11488
- Takahashi S, Kuzuyama T, Watanabe H, Seto H (1998) A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol

4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis.

Proc Natl Acad Sci U S A 95: 9879-9884

Wegener D, Steinecke P, Herget T, Petereit I, Philipp C, Schreier PH (1994)

Expression of a reporter gene is reduced by a ribozyme in transgenic plants. Mol

5 Gen Genet 1994 Nov 15;245(4):465-470

Zhou D, White RH. 1991. Early steps of isoprenoid biosynthesis in Escherichia coli.

Biochem J 273:627-634

Having now fully described the invention, it will be apparent to one of  
ordinary skill in the art that many changes and modifications can be made thereto  
10 without departing from the spirit or scope of the invention as set forth herein.

We claim:

1. A method of enhancing the production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, relative to an untransformed host cell, the method comprising
  - 5 inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having deoxyxylulose-5-phosphate (DXP) synthase enzyme activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and
  - expressing the heterologous nucleic acid sequence, thereby
  - 10 producing the protein.
2. The method of claim 1, wherein the isoprenoid is derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).
3. The method of claim 1, wherein the isoprenoid comprises at least one
- 15 member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.
4. The method of claim 1, wherein the isoprenoid comprises a carotenoid.
5. The method of claim 1, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *dxps*.
- 20 6. The method of claim 5, wherein the nucleic acid sequence for *dxps* encodes an amino acid sequence which is completely identical to SEQ ID NO: 2.
7. The method of claim 1, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
8. The method of claim 1, wherein the heterologous nucleic acid sequence

originates from a eukaryotic cell.

9. The method of claim 1, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

10. A method of modifying the production of isopentenyl pyrophosphate (IPP),  
5 dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, the method comprising

inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein which modifies deoxyxylulose-5-phosphate (DXP) synthase enzyme activity in the host cell, relative to an  
10 untransformed host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence, thereby producing the protein.

11. The method of claim 10, wherein the isoprenoid is derived from at least one  
15 member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).

12. The method of claim 10, wherein the isoprenoid comprises at least one member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.

20 13. The method of claim 10, wherein the isoprenoid comprises a carotenoid.

14. The method of claim 10, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

15. The method of claim 10, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

16. The method of claim 10, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
17. A method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having deoxyxylulose-5-phosphate (DXP) synthase enzyme activity, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.
18. The method of claim 17, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *dxps*.
19. The method of claim 18, wherein the nucleic acid sequence for *dxps* encodes an amino acid sequence which is completely identical to SEQ ID NO: 2.
20. The method of claim 17, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
21. The method of claim 17, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.
22. The method of claim 17, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
23. A method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein which modifies deoxyxylulose-5-phosphate (DXP) synthase enzyme activity in the host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

24. The method of claim 23, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
25. The method of claim 23, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.
- 5 26. The method of claim 23, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
27. A method of enhancing the production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, relative to an untransformed host cell, the method comprising
- 10 inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having glyceraldehyde-3-phosphate (GAP) dehydrogenase enzyme activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and
- expressing the heterologous nucleic acid sequence, thereby producing the
- 15 protein.
28. The method of claim 27, wherein the isoprenoid is derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).
29. The method of claim 27, wherein the isoprenoid comprises at least one
- 20 member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.
30. The method of claim 27, wherein the isoprenoid comprises a carotenoid.
31. The method of claim 27, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *gapd*.

32. The method of claim 31, wherein the nucleic acid sequence encodes an amino acid sequence which is completely identical to SEQ ID NO: 4.

33. The method of claim 27, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

5 34. The method of claim 27, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

35. The method of claim 27, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

10 36. A method of modifying the production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, the method comprising

inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein which modifies glyceraldehyde-3-phosphate (GAP) dehydrogenase enzyme activity in the host cell, relative to an untransformed  
15 host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence, thereby producing the protein.

20 37. The method of claim 36, wherein the isoprenoid is derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).

38. The method of claim 36, wherein the isoprenoid comprises at least one member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.

25 39. The method of claim 36, wherein the isoprenoid comprises a carotenoid.



40. The method of claim 36, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
41. The method of claim 36, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.
- 5 42. The method of claim 36, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
43. A method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having glyceraldehyde-3-phosphate (GAP) dehydrogenase enzyme activity, the method comprising inserting into the host cell  
10 a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.
44. The method of claim 43, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *gapd*.
- 15 45. The method of claim 44, wherein the nucleic acid sequence encodes an amino acid sequence which is completely identical to SEQ ID NO: 4.
46. The method of claim 43, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
47. The method of claim 43, wherein the heterologous nucleic acid sequence  
20 originates from a eukaryotic cell.
48. The method of claim 43, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
49. A method of expressing, in a host cell, a heterologous nucleic acid sequence

which encodes for a protein which modifies glyceraldehyde-3-phosphate (GAP) dehydrogenase enzyme activity in the host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

50. The method of claim 49, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

51. The method of claim 49, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

52. The method of claim 49, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

53. A method of enhancing the production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, relative to an untransformed host cell, the method comprising

inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having LYTB activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and  
expressing the heterologous nucleic acid sequence, thereby producing the protein.

54. The method of claim 53, wherein the isoprenoid is derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).

55. The method of claim 53, wherein the isoprenoid comprises at least one member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.

56. The method of claim 53, wherein the isoprenoid comprises a carotenoid.
57. The method of claim 53, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *lytB*.
58. The method of claim 57, wherein the nucleic acid sequence encodes an amino acid sequence which is completely identical to SEQ ID NO: 6.
59. The method of claim 53, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
60. The method of claim 53, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.
61. The method of claim 53, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.
62. A method of modifying the production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or an isoprenoid in a host cell, the method comprising
- inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein which modifies LYT<sub>B</sub> activity in the host cell, relative to an untransformed host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and
- expressing the heterologous nucleic acid sequence, thereby producing the protein.
63. The method of claim 62, wherein the isoprenoid is derived from at least one member selected from the group consisting of geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP).
64. The method of claim 62, wherein the isoprenoid comprises at least one

member selected from the group consisting of an essential oil, a fragrance, a rubber, a carotenoid and a therapeutic compound.

65. The method of claim 62, wherein the isoprenoid comprises a carotenoid.

5 66. The method of claim 62, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

67. The method of claim 62, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

68. The method of claim 62, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

10 69. A method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having LYT<sub>B</sub> activity, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

15 70. The method of claim 69, wherein the heterologous nucleic acid sequence comprises a nucleic acid sequence for *lytB*.

71. The method of claim 69, wherein the nucleic acid sequence encodes an amino acid sequence which is completely identical to SEQ ID NO: 6.

20 72. The method of claim 69, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

73. The method of claim 69, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

74. The method of claim 69, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

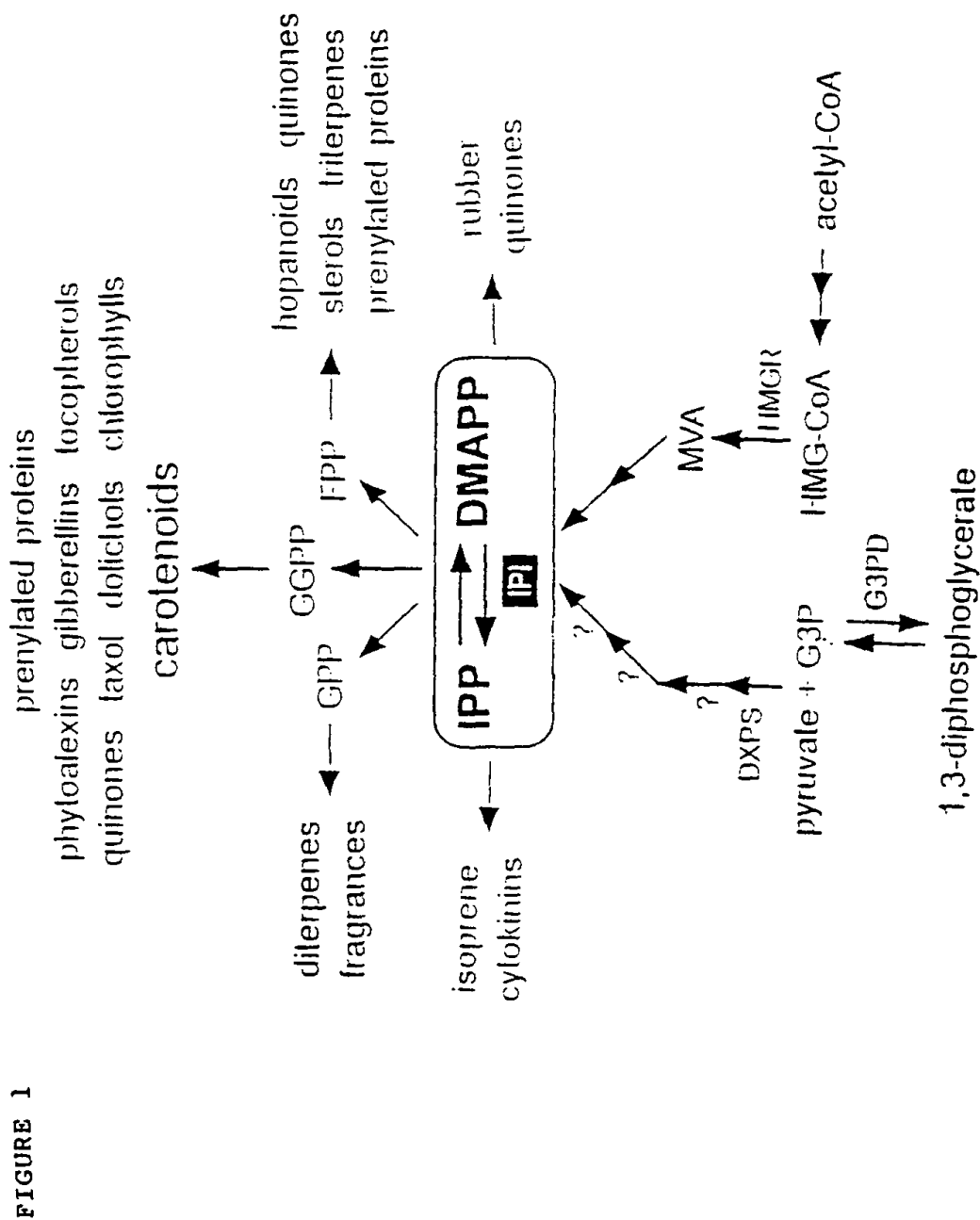
75. A method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein which modifies LYTB activity in the host cell, relative  
5 to an untransformed host cell, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter, and expressing the heterologous nucleic acid sequence.

76. The method of claim 75, wherein the host cell is selected from the group  
10 consisting of a bacterial cell, an algal cell and a plant cell.

77. The method of claim 75, wherein the heterologous nucleic acid sequence originates from a eukaryotic cell.

78. The method of claim 75, wherein the heterologous nucleic acid sequence originates from a procaryotic cell.

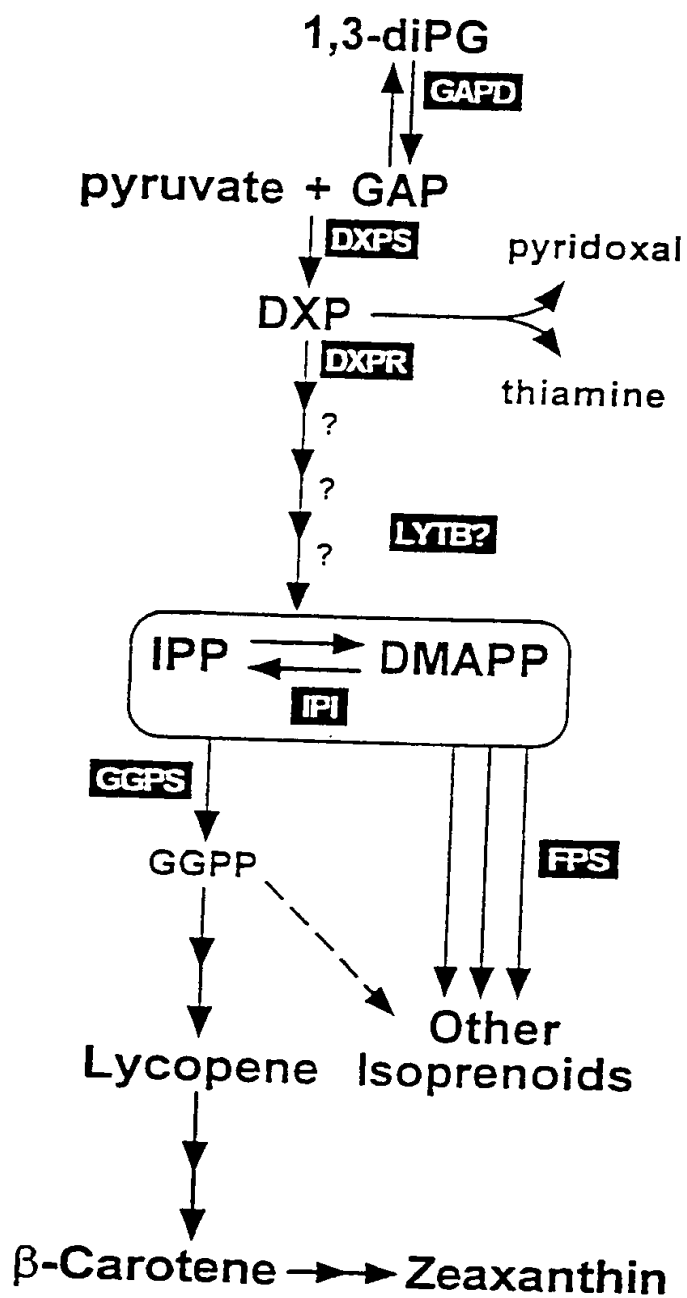
1/13



SUBSTITUTE SHEET (RULE 26)

2/13

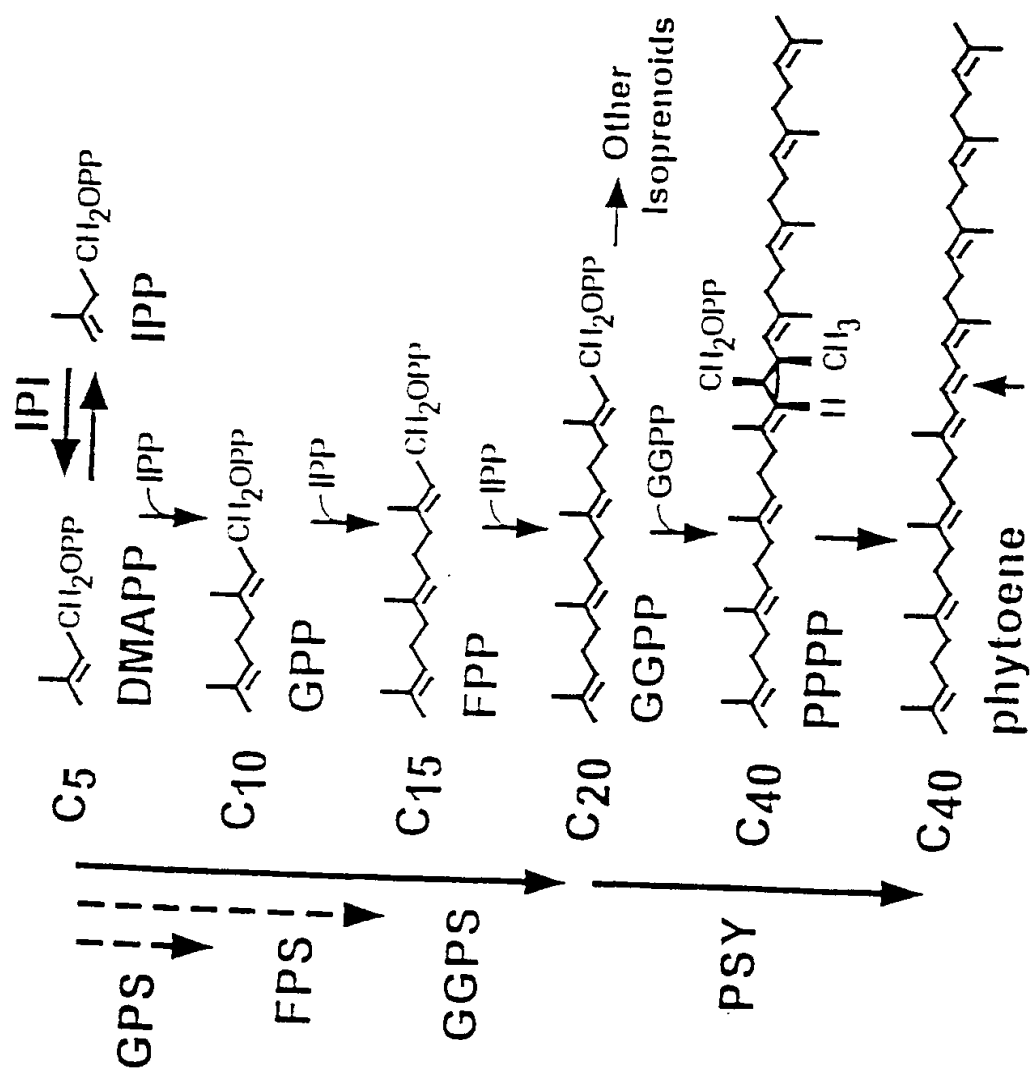
FIGURE 2



SUBSTITUTE SHEET (RULE 26)

3/13

FIGURE 3





4/13

FIGURE 4

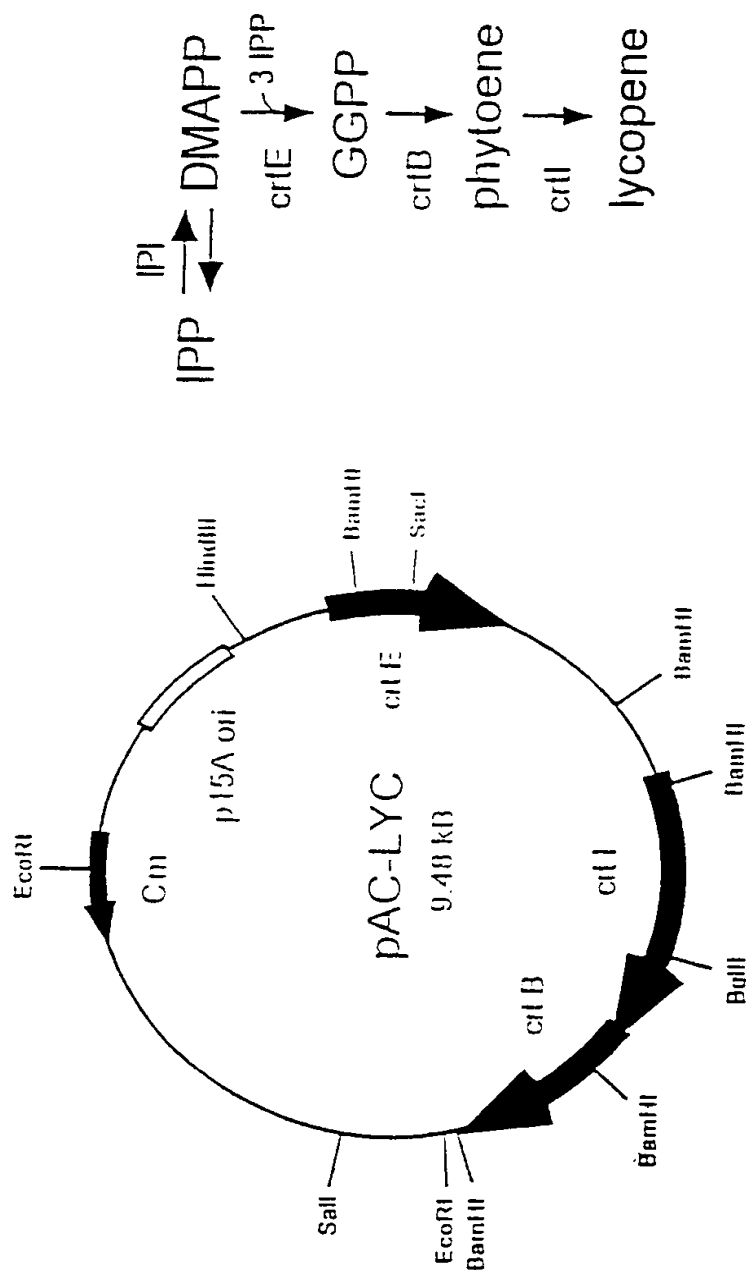


Figure 5A: SEQ ID NO: 1

5/13

*Tagetes erecta* (marigold) cDNA for DXP synthase

```
1  ggcacgagaa gcactctcat attctcttgt gtgttttctg aatcacacca
51  acaaatggct ttgtgtggtg ctttgaaggg tgggtttgtg cccatagctc
101 aaaatggtta cacttctcca agtttgctta atcgtcagc caatgcaatc
151 atgccatcaa acaagcgaaa gtttttggga atcgtagctg tttcaaaaga
201 acatgcaacc aatgaacatg aggacctgac aacaatggat aagactacat
251 caaccaccct caagtattct ggggataaac ccaaaactcc gatcttggat
301 acgattaatt atccaattca catgaaaaat ctttgctggtg aggagctcgt
351 aaaattagct gatgaactcc gagaagaaat tgtatatacg gtatcaaaga
401 caggcggtca cttgagttcg agcttggggg ttgttgaatt aacagtttca
451 ctccatcatg ttttcaacac tccagaagac aagatcatct gggacgttgg
501 tcatcaggct tatccacata aaatattgac cggagaaggg tcgagaatga
551 gaacgatttcg acaaacgttt ggacttgcag gcttccctaa gagagacgag
601 agtaatcatg atgcttttgg tgcaggccat agctctacta gcatttctgc
651 aggttttagga atggctgttg gtagagactt gttaggaaag aacaatcatg
701 taattgcagt gatcgggtgat ggagcgatga ccgcaggaca agcatatgag
751 gccatgaata atgcaggtta cctagattct aatctcatca tagtcttaaa
801 tgacaacagg cagttctctt tgcccacagc cactattgat ggtcctgcgc
851 cgctgttggt cgcactcagc cgatccctca caaggcttca aacaagccaa
901 aagtttcgct aactacgtga agccgcaaag gaagtaacaa agcaattagg
951 ggataaaaca catgaagtgt cggcaaaaat ggactctcta gtgaaaggaa
1001 tgggttggtg tcaagggtct tctatgttgg aagagcttgg cctctattat
1051 gttggtccgg tagatggaca caatcttgaa gatcttgtct atgttcttga
1101 taagattaag tccatgacgg ctccgggacc agtcctagtc cacattgtta
1151 ccgaaaaagg caaagggtac ccgcccgtg aagttgctgc agataaaatg
1201 catggtgttg tgaaatttga tactcaaacc ggcaaacaaa agaaaaataa
1251 gacaaaaaca ctctcatata cccaatattt cgtggactct ttgggtgctg
1301 aagcgaaaga ggatgacaag atagtcgcga ttcatgcggc tatgggcgga
1351 ggcacaggcc ttaacacatt ccaaaaagag ttccctgcac gttgttttga
1401 tgttggcata gctgagcaac acgcaatcac ctttgctgca ggtttagcaa
1451 cagaagggtt gaaaccattt tgtgctattt attcatcttt cctacaaaga
1501 ggttatgata aagtggttca tgatgttgat cttcaaaagc ttctgttag
1551 atttgctatg gatagggtgt gtttggtagg agcagatggt ccaactcact
1601 gtggtgcatt tgatacaaca ttcattggct gtttaccaaa catggtggtt
1651 atggctcctt catgcgaggc agagctcatg aacatggtgg ccacagctgt
1701 agccatcgat gaccgtccta gctgttttag gtacccaaga ggaaatggaa
1751 taggctccat actaccagca aataacaaag gaaccttaat cgaggttgga
1801 actggaaggg tgattaaaga gggaaataga gtggccttat tgggatatgg
```

SUBSTITUTE SHEET (RULE 26)

FIGURE 5A (cont.)

6/13

```
1851 aactatagtc caaagttgct tagcagcaag tgaagttctt aaaaaaattg
1901 gcatctctgt gaccgtggcg gatgcacgat tctgcaagcc gcttgatgga
1951 aattttaatca aacaactagc aaacgagcat gaggtcctca tcacagttga
2001 agaaggctct ataggaggat ttagctctca cgtctctcat ttcttggcct
2051 taaatggatt acttgatgga caccttaagt ggagagcaat gatgttgect
2101 gatcgatata ttgaacatgg ggctcaatct gaccagattg aggaggctgg
2151 tcttagctca aagcataattg cagcaacagt tttatcattg attgggggca
2201 gcaaagagac actacatgca ctcaacgtat aaattgaaat atgtcatgct
2251 agatactata aaagacaatg ctttgctgga ctcttttgat catatatata
2301 tatatatata tatatatata tatatatatt gtagctttcg atcatataca
2351 tattgtactt ttcttgtaac atacatggct gcgttgataa ttgttgacat
2401 aatctcgctc
```

SUBSTITUTE SHEET (RULE 26)

Figure 5B: SEQ ID NO: 2

7/13

PREDICTED AMINO ACID SEQUENCE of *Tagetes erecta* (marigold) cDNA for  
DXP synthase

TRANSLATION from: 55 to: 2232      Length: 726

```
1  MALCGALKGG FVPIAQNGYT SSSLLNPSAN AIMPSNKRKF LGIVAVSKEH
51  ATNEHEDLTT MDKTTSTTLK YSGDKPKTPI LDTINYPIHM KNLCVEELVK
101 LADELREEIV YTVSKTGGHL SSSLGVVVELT VSLHHVFNTF EDKIIWDVGH
151 QAYPHKILTG RRSRMRTIRQ TFGLAGFPKR DESNHDAFGA GHSSTSISAG
201 LGMAVGRDLL GKNNHVIAMI GDGAMTAGQA YEAMNNAGYL DSNLIIVLND
251 NRQFSLPTAT IDGPAPPVGA LSRSLTRLQT SQKFRQLREA AKEVTKQLGD
301 KTHEVAAKMD SLVKGMVGGQ GASMFEELGL YYVGPVDGHN LEDLVYVFDK
351 IKSMTAPGPV LVHIVTEK GK GYPPAEVAAD KMHGVVKFDT QTGKQKKNKT
401 KTLSTYQYFV DSLVAEAKED DKIVAIHAAM GGGTGLNTFQ KEFPARCFDV
451 GIAEQHAITF AAGLATEGLK PFCAIYSSFL QRGYDQVVHD VDLQKLPVRF
501 AMDRAGLVGA DGPTHCGAFD TTFMACLPNM VVMAPSCEAE LMNMVATAVA
551 IDDRPSCFRY PRGNGIGSIL PANNKGTLE VGTGRVIKEG NRVALLGYGT
601 IVQSCLAASE VLKKIGISVT VADARFCKPL DGNLIKQLAN EHEVLITVEE
651 GSIGGFSSHV SHFLALNGLL DGHLKWRAMM LPDRYIEHGA QSDQIEEAGL
701 SSKHIAATVL SLIGGSKETL HALNV*
```

SUBSTITUTE SHEET (RULE 26)

Figure 6A: SEQ ID NO: 3

8/13

*Arabidopsis thaliana* cDNA for a chloroplasts isoform of glyceraldehyde-3-phosphate (GAP) dehydrogenase.

```

1  ctgtcccca gggtttcact gaattctcag gattgcgaag ctccctctgct
51  tctcttccct tcggcaagaa actttcttcc gatgagttcg tttccatcgt
101 ctccctccag acttctgcaa tgggaagcag tgggtggatac aggaaagggtg
151 tgactgaggc caagcttaag gtggccatta atggattcgg taggatcggg
201 aggaacttcc tgagatgttg gcatggtcgc aaggactctc ctcttgatat
251 cattgccatt aatgacactg gtggcgtaaa gcaggcttcg catttactta
301 aatacgactc tactctcgga atctttgatg ctgatgtcaa accttctgga
351 gagactgcaa tctctgttga tggaaagatc atccaagttg tctctaaccg
401 aaaccctctt cttctccctt ggaaggagct aggaattgac attgtcatcg
451 aaggaaccgg agtgtttggt gatagagaag gtgcagggaa acacattgaa
501 gctgggtgca agaaggatat cttactgct ccaggcaaag gagatattcc
551 aacttatgtc gttgggtgtc atgcagatgc ttacagtcac gatgaacctc
601 tcatcagcaa tgcattcttc actaccaact gtctttgctcc ctttctcaaa
651 gttcttgacc agaaattcgg taccataaag ggtacaatga cgactactca
701 ctcttacacc ggtgaccaga ggttgctaga cgcgagtcac cgtgatctaa
751 ggagagcaag agcagctgct ttgaacattg ttccctacttc tacaggagca
801 gctaaagctg tggctctctg gctccctaac ctcaaaggaa aactcaacgg
851 gatcgctctc cgtgtacca caccaaacgt atcagtgggt gatctcgttg
901 tgcaggcttc aaagaagaca ttgctgagg aagtcaacgc tgccttcaga
951 gattctgcag agaaagagct taaaggcata ctcgatgtct gcgatgagcc
1001 actagtgtcc gttgatttca gatgctcaga tttttcaacg accattgatt
1051 catcactcac tatgggttat ggagatgata tggttaagggt gattgcttgg
1101 tatgataatg aatgggggtc ctacagaga gttgttgact tggctgacat
1151 tgttgccaac aactggaagt gatttcacaa aactttgctt catgttttcc
1201 catttttttc tctttctctt ttttacatta tgggtctcaa aatatcgggtg
1251 aagactgtat a

```

SUBSTITUTE SHEET (RULE 26)

Figure 6B: SEQ ID NO: 4

9/13

Predicted amino acid sequence of *Arabidopsis thaliana* cDNA for a chloroplasts isoform of glyceraldehyde-3-phosphate (GAP) dehydrogenase.

```
1  VPKGTFTEFSG LRSSSASLPF GKQLSSDEFV SIVSFQTSAM GSSGGYRKGV
51  TEAKLKVAIN GFGRIGRNFL RCWHGRKDSP LDIIAINDTG GVKQASHLLK
101 YDSTLGIFDA DVKPSGETAI SVDGKIIQVV SNRNPSLLPW KELGIDIVIE
151 GTGVFVDREG AGKHIEAGAK KVIITAPGKG DIPTYVVGVN ADAYSHDEPI
201 ISNASCTTNC LAPFVKVLDQ KFGIIKGTMT TTHSYTGDQR LLDASHRDLR
251 RARAAALNIV PTSTGAAKAV ALVLPNLK GK LGIALRVPT PNVSVVDLVV
301 QVSKKTFAEE VNAAFRDSAE KELKGILDVC DEPLVSVD FR CSDFSTTIDS
351 SLTMVMGDDM VKVIAWYDNE WGYSQRVVDL ADIVANNWK* FHQTLFHVFP
401 FFSLSLFYIM VLKISVKTV
```

SUBSTITUTE SHEET (RULE 26)

Figure 7A: SEQ ID NO: 5  
*Adonis palaestina* cDNA  
Length: 1619 June 4, 1998

10/13

```
1  ccttcgtctc tgcacttctt tcttccatgg cgacttctct tcaattctgc
51  cgttttctcca ccccatctga tctctccttc cctgaaaccc gctcttccac
101  tcgcctctac cgatccaaga aaccattttc cgtccgttgt cacagtgaag
151  gaccttcagg ctccctcctcc acggcagtgg aatccgaatt cgatgccaag
201  tctttccgcc acaaccttac aagaagcaag aactacaacc gaagagggtt
251  cggccacaaa gatgagactc ttgagcttat gaacagcgag tatacaagt
301  atgttattaa gaagttgaag gagaatggaa acgagtacag ttggggacct
351  gttacagtga aattggctga gtcgtatggg ttctgttggg gtgtggaaag
401  agctgttcag attgcttatg aagctagaaa acagtttccc gatgagaaga
451  tttggattac taatgagatt atccataatc caactgttaa caagagatta
501  gaagagatgg aagttaaaga gattccagtt ggggatggga agaaacattt
551  tgatgttggt gccaaagggtg acgttgtgat ttggccggct tttggagctg
601  ccgtaagtga aatgttgact ttgagcgaga aaaatgtcca gatagtcgac
651  acaacttgcc catgggtatc taagggtgtg aattccgttg agaagcataa
701  gaaggagag taccatcaa ttattcatgg taaatatcc cactaggaaa
751  ctattgcaac tgcattcttt gctgggaagt atatcattgt gaagaacatg
801  gatgaggcaa tgtatgtgtg cgactacatt cttggaggtg agcttaatgg
851  atccagctca gataaacagg cacttttggg gaaattcaaa tacgcaattt
901  ccgaaggatt tgatccagac acagatctta ttaaaacagg cattgcaaac
951  caaactacaa tgttgaaggg agaaacagaa gatattggta aacttctcga
1001 aaagactatg atgcgcaagt atggagttga gaatattaar gaccacttca
1051 taagctttta cacaatatgt gatgctacac aagagagaca agacgccatg
1101 ttttaagttg tagaggagaa agtggatctc atattagtag ttggtggatg
1151 gaattcaagt aacacctcac atctacaaga aatttctgag ttacgtggca
1201 ttcttcttta ctggattgac agtgagacga gaataggacc aggaaacaag
1251 attagccaca aattgaatca tggtgagcta gttgagaccg aaaactggct
1301 cccagagggt cctgttacaa ttggtgtgac atcaggtgct tctaccccag
1351 ataaggctgt tgaagatgct cttgtaaagg tgttcgacat caagcgcgaa
1401 caattgttgc agctagcata gagatcagtg ttctgcagag atcaaagaac
1451 caccactgtt caatatgtgt gattgtatat gtatcatgga gaacacaggg
1501 atatagaaac tgtaatacct tgttatgtta tactatatgg atgtccttta
1551 ccataatgat gtaaatTTTT agttccatgc aacaatattt tcttgaatga
1601 aaaaaaaaaa aaaaaaaaaa
```

SUBSTITUTE SHEET (RULE 26)

Figure 7B: SEQ ID NO: 6      11/13

Predicted amino acid sequence of Adonis palaestina LYTB

Length: 465 Translate 27-1421

```
1  MATSLQFCRF STPSDLSFPE TRSSTRLYRS KKPFSVRCHS EGPSGSSSTA
51  VESEFDAKSF RHNLTRSKNY NRRGFGHKDE TLELMNSEYT SDVIKKLKEN
101 GNEYSWGPVT VKLAESYGFC WGVERAVQIA YEARKQFPDE KIWITNEIIH
151 NPTVNRLEE MEVKEIPVGD GKXHFVVAK GDVVILPAFG AAVSEMLTSL
201 EKNVQIVDTT CPWVSKVWNS VEKHKKG EYT SIIHGKYSHE ETIATASFAG
251 KYIIVKNMDE AMYVCDYILG GELNGSSSDK QALLEKFKYA ISEGFDPD TD
301 LIKTGIANQT TMLKGETEDI GKLEKTMMR KYGVENINDH FISFNTICDA
351 TQERQDAMFK LVEEKVDLIL VVGWNSNT SHLQEISELR GIPSYWIDSE
401 TRIGPGNKIS HKLNHGELVE TENWLPEGPV TIGVTSGAST PDKAVEDALV
451 KVFDIKREQL LQLA*
```

**SUBSTITUTE SHEET (RULE 26)**



12/13

FIGURE 8

```

Ap : MATSLQFCRFSTPSDLSEFPETRSSTRLYRSKKPF SVRCHSEGPSSGSSSTA 50
Sy : -----MDTKAFKRSLHHS DNYHRKGFGHGGEVMGVMNTEYQSHLIQEI RQN
Ec : -----

Ap : VESEFDAKSEFRHNLTRSKNYNRRGFGHKDETLELMNSEYTS DVIKKLKEN 100
Sy : -----MDTKAFKRSLHHS DNYHRKGFGHGGEVMGVMNTEYQSHLIQEI RQN
Ec : -----

Ap : GNEYSWGPVTVKLAESY GFCWGVERAVQIAYEARKQEPDEKIWITNEITH 150
Sy : NYRLERGDVTILLAEAF GFCWGVERAVAMAYETROHEFGDRLWITNEITH
Ec : -----MOTLLANPR GFCAGVDRAISTVENAL-AIYGAPTYVRHEVVH

Ap : NPTVNKRLEEMEYKEIPVGDGKKHFDV VAKG D VVILPAEFGAAVSEMLTIS 200
Sy : NPSVNQRIRREMEYNFIDV VNGEKDFSGVAKG D VVILPAEFGASVEEMQLLN
Ec : NRYVVDSLRERGAIFIE -----QISEVPD GAILHFSAHGVSQAVRNEAK

Ap : EKNVQIVDTTCPPWVSKVWNSVEKHKKG EYTSI IHGKY SHEETIATASFAG 250
Sy : DRECHIVDTTCPPWVSKVWNSVEKHKKKEHTSI IHGKYNHEETIATSSFAG
Ec : SRDLTVFDATCPLVTRVHMEVARASRRGEE SILIGHAGHPPEVEGTM-----

Ap : KYLIVKNMDEAMYVCDYILGGEINGSSSDKQALLEKEKYAISEGFDPD TD 300
Sy : TYLLFVLNMAEAQKVCDYILHG -----GDRQEFLDYEANAH SAGEDPDQD
Ec : -----GQYSNPEGG-MYLVE SPDDVWKLTVKNEE-

Ap : EIRTGIANOTTMLKGETEDIGKILLEKTMMRKYGVENINDHEISENTICDA 350
Sy : TVRIGVANOTTMLKSETEMTGKLFERKTLLOKYGPTELNNHFMSENTICDA
Ec : --KLS EMTOTTLSVDDTSDVIDALRRKREP KIVGPRK-----DDICYA

Ap : TQERQDAMFKLVEEKVDLILVVGWNSSNTSHIQEISELRGIPSYWIDSE 400
Sy : TQERQDAMFDLVEEDLSLMVVGGENSSNTTHLQELAVEHGIPSVHIDSG
Ec : ITNRQEAVALAE-QAEVVLVVGSKNSSNSNRLAE LAQRMGKRAFLIDDA

Ap : TRIGPGNKNISHKLNHGELVETENWLPEGPVTIGVTS GASTPDKAVEDALV 450
Sy : DRIGPGNRVERKPLGKDLEVLEPWLPAGKITVGVTS GASTPDKVVEEVMK
Ec : KDIQ-----EEWV-KEVKCYGVTAGASAPDILVQNVVA

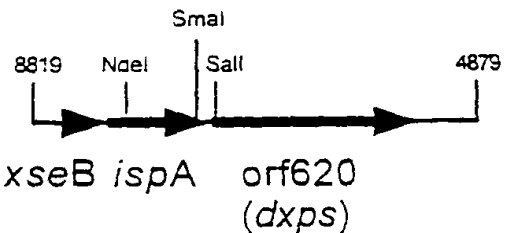
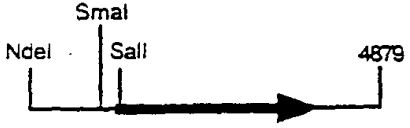


Ap : KVFDIKREQ-LLQLA-----: 464
Sy : KILAIKEAQPVLEIAG-----: 406
Ec : RLQQQLGGGEAIPLEGREENIVFEVPKELRVDIRVD: 316

```

SUBSTITUTE SHEET (RULE 26)

13/13

FIGURE 9

Plasmid	Insert	Lycopene Accumulation (as % of control)
pSK-	Control (empty vector)	100
pEc3.9	 <p>8819 NdeI SmaI SalI 4879</p> <p><i>xseB</i> <i>ispA</i> <i>orf620</i> (<i>dxps</i>)</p>	84 ± 1
pEc3.1	 <p>NdeI SmaI SalI 4879</p>	158 ± 9
pEc2.4	 <p>SmaI SalI 4879</p>	131 ± 3
pEc1.6	 <p>8819 NdeI SalI</p>	60 ± 5

SUBSTITUTE SHEET (RULE 26)

## INTERNATIONAL SEARCH REPORT

National application No.  
PCT/US99/07041

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/02, 9/04, 1/20, 15/00; C07H 21/04.

US CL : 435/189, 190, 252.3, 320.1, 468; 536/23.2, 23.6.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/189, 190, 252.3, 320.1, 468; 536/23.2, 23.6.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,539,093 A (FITZMAURICE et al.) 23 July 1996. See, the entire patent.	1-78
A	US 5,618,988 A (HAUPTMANN et al.) 08 April 1997. See the entire patent.	1-78
A	US 5,750,865 A (BIRD et al.) 12 May 1998. See the entire patent.	1-78

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

24 AUGUST 1999

Date of mailing of the international search report

19 OCT 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

TEKCHAND SAIDHA

Telephone No. (703) 308-0196

JOYCE BRIDGERS  
PARALEGAL SPECIALIST  
CHEMICAL MATRIX  
*gab*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/07041

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN files : Medline, Caplus, Biosis, Agricola, embase & Registry. Search terms : Carotenoid biosynthesis, deoxyxylulose 5-phosphate synthase, IPP, DMAPP, LYTB - in different combinations.

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING**

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-35, drawn to a method of enhancing production of isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP) and/or isoprenoid using a nucleic acid encoding deoxyxylulose 5-phosphate synthase (DXP) synthase.

Group II, claim(s) 36-52, drawn to a method of modifying the production of IPP, DMAPP and/or isoprenoid using a nucleic acid encoding a protein that modifies glyceraldehyde-3-phosphate dehydrogenase.

Group III, claim(s) 53-74, drawn to a method of enhancing the production of IPP, DMAPP and/or isoprenoid using a nucleic acid encoding a protein having LYTB activity.

Group IV, claims 75-78, drawn to a method of expression using a nucleic acid encoding a protein which modifies LYTB activity.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I has the special technical feature of a method using a nucleic acid encoding DXP-synthase, which Groups II-IV do not share. Group II has the special technical feature of a method using a nucleic acid encoding a protein which modifies G-3-P dehydrogenase, which Groups I and III-IV do not share. Group III has the special technical feature of a method using a nucleic acid encoding a protein having LYTB activity, which Groups I-II and IV do not share. Group IV has the special technical feature of a method using a nucleic acid encoding a protein which modifies LYTB activity, which Groups I-III do not share.

